L'épidémiologie ou la science de l'estimation du risque en santé publique

Arnaud Fontanet

Leçons inaugurales
du Collège de France

COLLÈGE
DE FRANCE
1530

L'épidémiologie ou la science de l'estimation du risque en santé publique

Leçon inaugurale prononcée le jeudi 31 janvier 2019

Arnaud Fontanet

DOI : 10.4000/books.cdf.8263
Éditeur : Collège de France
Lieu d'édition : Paris
Année d'édition : 2019
Date de mise en ligne : 31 janvier 2020
Collection : Leçons inaugurales
ISBN électronique : 9782722605336

http://books.openedition.org

Édition imprimée
Date de publication : 20 novembre 2019
ISBN : 9782213713144

Référence électronique
FONTANET, Arnaud. *L'épidémiologie ou la science de l'estimation du risque en santé publique : Leçon inaugurale prononcée le jeudi 31 janvier 2019.* Nouvelle édition [en ligne]. Paris : Collège de France, 2019 (généré le 03 février 2020). Disponible sur Internet : <http://books.openedition.org/cdf/8263>. ISBN : 9782722605336. DOI : 10.4000/books.cdf.8263.

Ce document a été généré automatiquement le 3 février 2020.

© Collège de France, 2019
Conditions d'utilisation :
http://www.openedition.org/6540

Peut-on évaluer le risque d'avoir une maladie en lien avec nos gènes, nos comportements ou notre environnement ? L'épidémiologie, qui étudie la répartition et les déterminants des maladies dans la population, permet de répondre à ces questions. Le développement fulgurant de cette science, dans la seconde moitié du xx^{e} siècle, a permis de faire des progrès considérables dans l'identification des facteurs de risque des maladies cardiovasculaires et de nombreux cancers. Les progrès du séquençage du génome humain au début du xxi^{e} siècle et, d'une façon plus générale, l'accès aux données massives (*big data*), ont révolutionné la recherche en épidémiologie tout en lui imposant de nouveaux défis scientifiques et éthiques.

SOMMAIRE

Préface
Philippe Sansonetti

L'épidémiologie ou la science de l'estimation du risque en santé publique
Leçon inaugurale prononcée au Collège de France le jeudi 31 janvier 2019
Arnaud Fontanet

Épidémiologie descriptive : surveillance des maladies et modélisation
Épidémiologie analytique : étude des déterminants des maladies
Grandes contributions de l'épidémiologie à la santé publique
La période du doute en épidémiologie
L'avenir de l'épidémiologie

NOTE DE L'ÉDITEUR

La chaire annuelle Santé publique a été créée en partenariat avec l'agence nationale Santé publique France.

Cet ouvrage a été réalisé avec la chaîne d'édition structurée XML-TEI Métopes développée par le pôle Document numérique de la Maison de la recherche en sciences humaines (MRSH) de l'université Caen-Normandie.

Préface

Philippe Sansonetti

1 C’est un honneur et un plaisir de présenter le professeur Arnaud Fontanet pour sa leçon inaugurale. Il est le premier titulaire de la nouvelle chaire annuelle Santé publique du Collège de France.

2 Je tiens d’abord à remercier nos collègues de Santé publique France, en particulier François Bourdillon, son directeur général, pour leur réponse enthousiaste à notre demande de considérer le mécénat de cette chaire.

3 Il est important que le Collège de France s’implique dans l’enseignement de la Santé publique. Notre dernière chaire d’Épidémiologie remonte à Hyacinthe Vincent, de 1925 à 1936, encore que le sujet ait été abordé plus récemment dans la période d’existence de la chaire annuelle Savoir contre pauvreté par Esther Duflo[1] et Peter Piot[2].

4 Domaine passionnant et complexe, dont les déterminants et les exigences sont souvent mal connus ou mal compris de nos concitoyens, la santé publique justifie un immense effort pédagogique, surtout pour ce qui touche à la prévention.

5 En des temps où la rationalité scientifique n’a pas toujours bonne presse, le partenariat entre le Collège de France et Santé publique France est déterminant et a, j’en suis certain, un grand avenir.

6 La santé publique aborde le sujet de la santé de manière collective, dans toutes ses dimensions, avec l’objectif d’améliorer l’état de santé de la population. Elle fait appel à de nombreuses sciences : l’épidémiologie, la démographie, la socio-anthropologie, l’écologie et l’économie de la santé. La matière étant riche et protéiforme, il semblait donc impossible d’imaginer qu’un seul savant puisse en couvrir les diverses facettes. C’est pourquoi nous avons créé une chaire annuelle et décidé de consacrer la première année d’enseignement à l’épidémiologie.

7 Pour définir celle-ci, citons Alain-Jacques Valleron : « C’est la science qui étudie les variations de fréquences des maladies dans les groupes humains et recherche les déterminants de ces variations. Elle vise en particulier à la recherche des causes des maladies et de l’amélioration de leurs traitements et moyens de prévention[3]. » L’épidémiologie est donc descriptive, analytique, de plus en plus modélisatrice et

évaluative du risque, éventuellement interventionnelle. Elle est d'abord et avant tout multidisciplinaire.

8 Arnaud Fontanet est d'emblée apparu pouvoir apporter à la chaire annuelle de Santé publique excellence et multidisciplinarité. Pasteurien, spécialiste des maladies infectieuses émergentes, il rejoint ici notre illustre prédécesseur, Charles Nicolle, titulaire de la chaire de Médecine expérimentale de 1932 à 1936, et théoricien de l'émergence : « Il y aura des maladies infectieuses nouvelles [...]. C'est un fait fatal », écrivait-il dans son livre *Destin des maladies infectieuses*[4].

9 Arnaud Fontanet est docteur en médecine de l'université René Descartes et docteur en santé publique de la Harvard School of Public Health, professeur au Conservatoire national des arts et métiers (Cnam) et à l'Institut Pasteur, où il dirige l'unité Épidémiologie des infections émergentes, qui est un modèle de multidisciplinarité.

10 Il aurait pu poursuivre une confortable carrière de rhumatologue parisien s'il n'avait eu l'idée saugrenue, au cours de son internat, de rejoindre pour deux ans les équipes de Médecins sans frontières et Médecins du monde opérant en Thaïlande dans des camps de réfugiés khmers et birmans. Il y attrapa le virus de la santé publique, dont il n'a jamais guéri...

11 Il a initialement consacré son activité à la recherche sur le sida en Éthiopie, puis sur l'hépatite C, domaine dont il est devenu un expert internationalement reconnu grâce à ses travaux sur l'endémie qui continue de sévir en Égypte, où elle touche des millions d'habitants.

12 Plus récemment, Arnaud Fontanet et ses collaborateurs ont développé des travaux pionniers sur des pathogènes émergents comme le virus Zika. En 2016, ils ont publié deux articles rapportant, lors de l'épidémie de Polynésie, l'association entre l'infection Zika et le syndrome de Guillain-Barré ainsi qu'entre l'infection Zika et la microcéphalie. Ces travaux, marqués par leur réactivité et leur qualité méthodologique, ont permis de démontrer l'existence de ces deux complications majeures avant que n'apparaisse la grande épidémie qui a balayé le Brésil. Ils ont contribué à placer la France en tête de la recherche épidémiologique pour cette maladie.

13 Arnaud Fontanet est responsable du Centre de santé globale de l'Institut Pasteur, en appui de nombreux projets sur l'émergence se développant, en particulier, dans le Réseau international des Instituts Pasteur (RIIP).

14 Il enseigne également, a créé et codirige l'École Pasteur-Cnam de Santé publique dans le cadre de laquelle il dirige le master de sciences de santé publique – et je peux témoigner du succès professionnel des étudiants qui y sont formés. Enfin, comme il s'ennuie un peu, il dirige le cours « Épidémies globales, alerte-réponse » à Science-Po.

15 Ce n'est cependant pas uniquement l'homme d'action que reconnaît l'attribution de cette chaire, c'est aussi l'homme de réflexion sur l'épistémologie, l'avenir, les forces et les faiblesses de sa discipline, en particulier sur la santé publique au temps des *big data*.

NOTES

1. Pauvreté et développement dans le monde (2008-2009) : https://www.college-de-france.fr/site/esther-duflo/course-2009-01-19-17h00.htm.

2. L'épidémie du sida et la mondialisation des risques (2009-2010) : https://www.college-de-france.fr/site/peter-piot/_course.htm.

3. Alain-Jacques Valleron, « Brève histoire de l'épidémiologie avant le XXe siècle », *La Jaune et la Rouge*, n° 670, dossier *Épidémiologie : au service de la santé*, décembre 2011.

4. Charles Nicolle, *Destin des maladies infectieuses*, Paris, PUF, 1939, p. 226.

AUTEUR

PHILIPPE SANSONETTI

Professeur au Collège de France, titulaire de la chaire Microbiologie et maladies infectieuses

L’épidémiologie ou la science de l’estimation du risque en santé publique

Leçon inaugurale prononcée au Collège de France le jeudi 31 janvier 2019

Arnaud Fontanet

---

1 Monsieur l’Administrateur,
Mes chers collègues et confrères,
Chers amis, chère famille,
Mesdames et Messieurs,

2 Je souhaite débuter cette leçon inaugurale en remerciant très chaleureusement Philippe Sansonetti pour avoir initié avec Alain Fischer ce projet de création d’une chaire de Santé publique au Collège de France. Je souhaite également remercier Santé publique France d’avoir apporté son soutien à cette initiative. Je suis très heureux et très honoré d’avoir été choisi pour être le premier titulaire de cette chaire, et j’espère que je saurai vous faire partager ma passion pour la santé publique et pour l’épidémiologie au cours de cette leçon inaugurale.

3 La médecine est un art merveilleux, diront certains : n’y a-t-il pas de plus beau métier que celui de soigner, de guérir ? En tant que jeune médecin, j’en étais convaincu. En fin d’internat, j’ai souhaité vivre une expérience différente de celle des hôpitaux parisiens, et je suis donc parti avec Médecins sans frontières pour exercer dans les camps de réfugiés cambodgiens en Asie du Sud-Est. Alors que je constatais l’efficacité remarquable de la quinine dans le traitement du paludisme sévère et celle des antibiotiques dans le traitement des infections respiratoires et gastro-intestinales, je me demandais : « Comment ces enfants ont-ils été infectés ? N’aurions-nous pas pu, avec une meilleure organisation sanitaire des camps de réfugiés, prévenir, de façon collective ou individuelle, ces infections ? » Une fois ces questions posées, je basculais dans l’univers de la santé publique. Je continue de penser aujourd’hui que la médecine est un art merveilleux, et j’encourage les plus jeunes qui ont la passion de guérir à s’y engager, mais à ceux qui s’intéressent aux déterminants de la santé des populations et

aux actions qui permettent de l'améliorer, de très belles perspectives s'ouvrent avec l'étude de la santé publique. J'émettrai une mise en garde toutefois : la santé publique est une profession ingrate. Un malade vous remerciera de l'avoir guéri. N'attendez pas des remerciements des personnes auxquelles vous aurez évité la maladie.

4 Même s'il s'agit de la première chaire de Santé publique au Collège de France, la santé publique n'a pas été absente du Collège de France jusqu'à aujourd'hui. Les professeurs qui ont occupé les chaires de médecine dans les siècles passés ont traité des questions d'hygiène. Le plus emblématique de ces professeurs a sans doute été Jean-Noël Hallé, surnommé le « médecin des pauvres » pendant la Révolution française (il a également été le médecin de Napoléon). Il enseignait l'hygiène médicale au Collège de France, et il a défendu l'introduction du vaccin contre la variole. Pour un aperçu du contenu de la discipline, je conseille la lecture du traité d'Émile Beaugrand intitulé *L'Hygiène, ou l'art de conserver la santé* et publié en 1855[1]. Il y eut même une chaire d'Épidémiologie, occupée par Hyacinthe Vincent, qui a œuvré pour la mise au point d'un vaccin contre la fièvre typhoïde destiné aux soldats français pendant la Première Guerre mondiale. Plus récemment, la chaire Savoirs contre pauvreté, où se sont exprimés notamment Dominique Kerouedan et Peter Piot, spécialistes de santé mondiale, a beaucoup traité des questions de gouvernance et de volonté politique dans l'exercice de la santé publique. Mais qu'est-ce donc que la santé publique ? Charles-Edward Amory Winslow l'a définie en 1920 comme la science et l'art de prévenir la maladie, de prolonger la vie et de promouvoir la santé humaine à travers les efforts organisés et informés de la société, des organisations – publiques et privées –, des communautés et des individus[2]. De fait, la santé publique est par nature multidisciplinaire, faisant intervenir professions médicales, épidémiologistes, biostatisticiens, économistes de la santé, anthropologues, spécialistes des politiques publiques, de l'organisation des soins, mais également toxicologues, éthiciens, juristes, microbiologistes, etc. Bref, la liste est longue et non limitative. Parmi toutes ces disciplines, c'est l'épidémiologie que j'ai choisie, et dont je vais vous parler aujourd'hui.

5 Vous connaissez tous les méfaits du tabac, de l'alcool, d'une alimentation non équilibrée. Vous avez été sensibilisés au bénéfice du dépistage de l'hypertension artérielle, de certains cancers, ou de maladies infectieuses comme le sida. Savez-vous que ces recommandations sont issues d'études épidémiologiques ayant porté sur plusieurs centaines de milliers d'individus suivis pendant des dizaines d'années ? L'épidémiologie est la discipline scientifique qui étudie en population la répartition des maladies et leurs déterminants. L'épidémiologie est résolument orientée vers l'action, les résultats des études visant à contrôler les maladies. De fait, l'épidémiologie est au cœur de nombreux débats de société : glyphosate, chlordécone, perturbateurs endocriniens, particules fines, salmonelles, Ebola, autant de termes qui ont fait la une de nos journaux cette année. Dès lors, mon objectif au cours de cette leçon inaugurale est non seulement de vous familiariser avec les grands enjeux de l'épidémiologie, mais également de vous donner des clés pour mieux comprendre les résultats des études épidémiologiques et vous permettre d'avoir un regard critique sur leur interprétation.

6 La démarche scientifique suivie en épidémiologie repose sur l'estimation du risque – d'où le titre de cette leçon inaugurale – d'être ou de devenir malade, pour l'épidémiologie descriptive, centrée sur la surveillance des maladies et leur modélisation ; ou sur l'estimation de l'augmentation (ou de la diminution) du risque

d'être malade, associée à certaines caractéristiques comme nos gènes, nos comportements, notre environnement, pour l'épidémiologie analytique.

Épidémiologie descriptive : surveillance des maladies et modélisation

7 Nous allons débuter par l'épidémiologie descriptive, axée sur la surveillance des maladies en population. Celle-ci est centrée sur le recueil de données descriptives sur l'état de santé des populations.

8 Il n'y a pas de cours d'épidémiologie qui ne débute par une présentation des travaux de John Snow. Je ne vais pas déroger à cette règle. John Snow était un médecin anesthésiste du XIX^e^ siècle connu pour avoir introduit l'éther et le chloroforme en Angleterre. Il aurait d'ailleurs administré du chloroforme à la reine Victoria à la naissance du petit Léopold en 1853. D'un tempérament curieux, John Snow s'intéresse à de nombreux autres sujets. Dès 1849, il émet l'hypothèse que le choléra puisse être transmis par l'eau, et suspecte la présence de « poisons animaux » responsables de la transmission, à une époque où la théorie des miasmes était plus populaire. Il aura l'occasion de tester sa théorie en 1854, lors d'une épidémie de choléra à Londres. Grâce à une enquête minutieuse auprès des résidents du quartier, il établit une relation significative entre le nombre de cas de choléra à Soho et le réseau de distribution d'eau qui dessert le quartier. L'action qui l'a rendu célèbre est d'avoir fait enlever le bras de la pompe incriminée. Dans la vie courante, on parlerait de bon sens. En épidémiologie moderne, on parle d'expérience contrefactuelle : que se serait-il passé en l'absence d'exposition ? Point-clé dans la démarche d'inférence causale. Le nombre de cas de choléra a décliné après le retrait du bras de la pompe, mais les esprits chagrins noteront que le déclin du nombre de cas avait débuté avant l'intervention. La même année, la bactérie responsable du choléra, *Vibrio cholerae*, était identifiée par un microbiologiste italien, Filippo Pacini. Il faudra cependant attendre la « redécouverte » de la bactérie par Robert Koch en 1883 pour que son rôle dans le choléra soit unanimement reconnu.

9 Qui sont les John Snow du XXI^e^ siècle ? On les retrouve dans les agences de sécurité sanitaire, dont le rôle est la surveillance de la santé des populations et l'évaluation des risques. Il s'agit entre autres de Santé publique France, de l'ANSES, qui associe également un volet santé animale, ou de l'ANSM pour les risques liés aux médicaments. La surveillance joue aujourd'hui un rôle essentiel de suivi des états de santé sur le territoire national ou planétaire. Elle s'appuie sur des réseaux de cliniciens, de vétérinaires, des laboratoires, des registres, ou des enquêtes en population. Elle détecte des « signaux », par exemple un début d'épidémie, qui déclenchera l'intervention des équipes d'investigation. Elle permet d'estimer le fardeau de la maladie (*burden of disease* des Anglo-Saxons), afin d'adapter les moyens de prise en charge aux besoins de la population. Enfin, elle permet d'évaluer l'impact des actions de prévention et des actions curatives mises en place par les pouvoirs publics. Initialement dédiée aux maladies infectieuses, la surveillance s'est étendue aux maladies chroniques comme les maladies cardiovasculaires, les cancers ou les maladies neurodégénératives.

10 Je vous invite à lire le rapport intitulé *État de santé de la population en France*, publié en 2017 par Santé publique France et la Direction de la recherche, des études, de

l'évaluation et des statistiques[3] (Drees). Ce rapport reprend et étoffe le suivi des cent objectifs de la loi de 2004 relative à la politique de santé publique en France, et je tiens à saluer ici le travail de William Dab, professeur du Conservatoire national des arts et métiers (Cnam) qui a beaucoup contribué à l'élaboration de cette loi de santé publique alors qu'il était directeur général de la santé. Ce travail est essentiel, car il nous donne un véritable tableau de bord, permettant de suivre l'impact des politiques menées, et d'identifier les zones de faiblesse, les populations à risque, et donc de hiérarchiser nos actions.

11 Les nouvelles y sont plutôt bonnes, si l'on regarde les tendances sur le long terme. L'espérance de vie à la naissance est de 85 ans pour les femmes, deuxième rang européen derrière l'Espagne, et de 79 ans pour les hommes. Plus de la moitié des décès en France sont dus aux maladies cardiovasculaires et aux cancers, à peu près à part égale. À distribution d'âge constante, les taux de mortalité liée aux cardiopathies ischémiques, c'est-à-dire au syndrome coronarien aigu et à l'infarctus du myocarde, ont diminué de plus de 40 % entre 2000 et 2013, la France ayant le taux de mortalité par cardiopathie ischémique le plus bas d'Europe. La mortalité liée au cancer est en légère baisse, de 1,5 % en moyenne par an chez les hommes à 1 % par an chez les femmes au cours des quarante dernières années, mais une analyse cancer par cancer est plus pertinente et nous y reviendrons un peu plus loin dans cette leçon.

12 À titre d'exemple, regardons l'apport de la surveillance de la rougeole en France. La France a connu une épidémie importante en 2011, suivie de quelques soubresauts les années suivantes, avant une recrudescence notable en 2018, totalisant plus de 27 500 cas pendant les dix dernières années. L'incidence de la maladie est plus importante chez les patients de moins d'un an, qui n'ont pas encore reçu une vaccination complète. Mais les formes les plus graves sont constatées chez les plus âgés, chez qui la rougeole est physiologiquement plus sévère. Il est désolant de devoir déplorer plus de 1600 pneumopathies graves, 40 encéphalites et plus de 20 décès au cours des dix dernières années pour une maladie pour laquelle on dispose d'un vaccin sûr et efficace. En effet, la couverture vaccinale n'est que de 80 % à 2 ans, alors qu'elle devrait être de 95 % pour bloquer la circulation du virus, tant il est contagieux. Vous connaissez tous le débat en cours sur la vaccination anti-rougeoleuse. Il est essentiel de disposer de chiffres fiables pour estimer le coût humain d'une protection vaccinale incomplète, et la surveillance des maladies à déclaration obligatoire est la principale source de ces chiffres.

13 La surveillance des maladies chroniques comme le cancer est également riche d'informations. La bonne nouvelle est que les taux de mortalité standardisés sur l'âge des principaux cancers sont en baisse sur la période 1950-2010. Cette baisse peut être due à une diminution du nombre de nouveaux cas par an grâce notamment aux efforts de prévention, ou à une diminution de la létalité des cancers grâce à une meilleure prise en charge thérapeutique, ou les deux. Nous observons chez l'homme des baisses importantes, comme celles du cancer du poumon, du fait de la baisse de la consommation tabagique, des cancers oro-pharyngés, du fait de la baisse de la consommation d'alcool et de tabac, et du cancer de l'estomac, du fait de la baisse des infections à *Helicobacter pylori*. Un point de vigilance demeure cependant : l'augmentation de la mortalité par cancer du poumon chez les femmes, liée à la non-diminution du tabagisme féminin pendant les décennies récentes.

14 Il est des situations où les données n'existent pas encore, notamment lors d'une épidémie débutante, et où les pouvoirs publics comme la communauté scientifique veulent savoir ce que sera le devenir de l'épidémie. Ce fut le cas lors de l'épidémie de variants de la maladie de Creutzfeldt-Jakob au Royaume-Uni, connue également sous le nom de « maladie de la vache folle ». Les premiers cas humains ont été détectés au milieu des années 1990. Certains modèles, issus d'équipes de recherche réputées, avaient prédit jusqu'à 136 000 cas d'ici à 2020[4], prédiction évidemment très inquiétante pour les pouvoirs publics du fait de l'extrême gravité de la maladie, avec cependant beaucoup d'incertitudes sur les prédictions. Ces incertitudes reposaient en grande partie sur les hypothèses émises autour de la durée d'incubation de la maladie. Plus elle était longue, plus le risque d'une épidémie importante était élevé. Alain-Jacques Valleron et son équipe[5] ont repris ces estimations sur la base d'une constatation : l'âge médian des cas était peu élevé, 28 ans. Pour expliquer ce fait, ils ont envisagé deux scénarios : soit les sujets les plus jeunes (moins de 15 ans) sont beaucoup plus susceptibles à l'infection, soit l'incubation est plus longue chez les sujets âgés, et l'épidémie ne serait encore qu'à ses débuts. Les catégories d'âge des patients ne variant pas avec l'année d'apparition des cas, l'idée d'une augmentation de la durée d'incubation des cas avec l'âge à la contamination était écartée. C'est donc l'hypothèse d'une susceptibilité accrue à un âge jeune qui a été retenue pour la modélisation, et la durée d'incubation moyenne a pu ainsi être estimée à 16,7 ans. Sur cette base, le modèle a prédit que le nombre total de cas de variants de la maladie de Creutzfeldt-Jakob au Royaume-Uni serait de 205, avec une fourchette haute à 403, bien loin des 136 000 avancés par les précédents modèles, et très proche des 177 finalement observés.

15 La modélisation peut être également utile aux pouvoirs publics quand il s'agit de tester l'efficacité d'interventions. C'est d'ailleurs à partir de simulations basées sur des modèles mathématiques que le gouvernement britannique a choisi l'abattage dans les 24 heures du bétail en cas d'infection avérée dans une ferme, tout en étendant l'abattage aux fermes voisines dans les 24 à 48 heures pour venir à bout de l'épidémie de fièvre aphteuse en 2001.

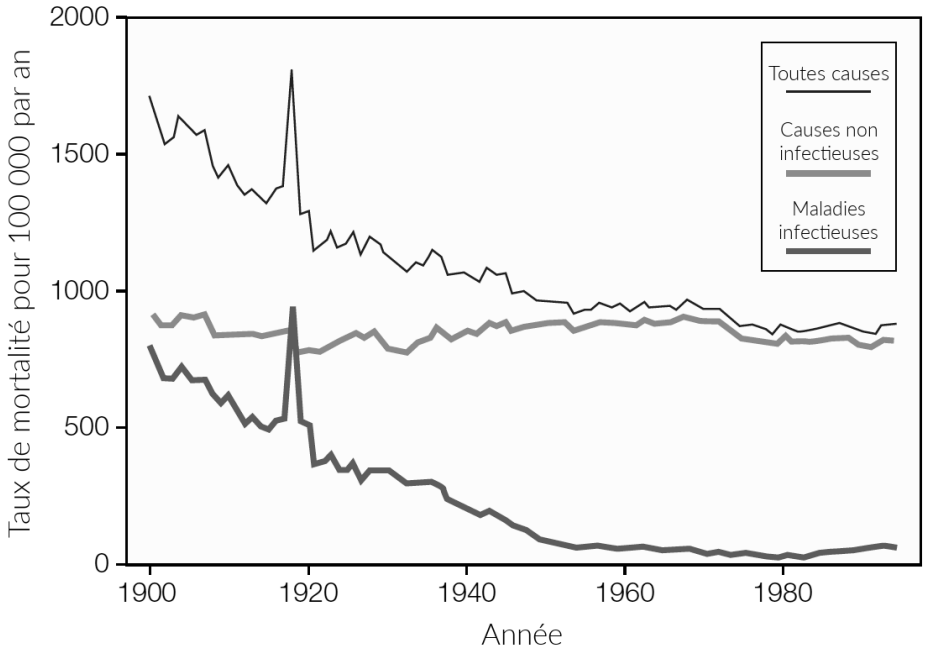
Épidémiologie analytique : étude des déterminants des maladies

16 L'autre grand versant de l'épidémiologie est dénommée « épidémiologie analytique ». Son objectif est d'identifier les déterminants des maladies. Il peut s'agir de nos gènes, mais également de ce qu'on appelle en épidémiologie nos « expositions », à savoir nos comportements (alcool, tabac, alimentation, etc.), la pollution atmosphérique, les médicaments que nous prenons, les agents infectieux présents dans notre environnement, parfois transmis par des vecteurs comme les moustiques, etc. Par analogie au génome, on parle d'ailleurs aujourd'hui d'exposome pour l'ensemble des expositions non génétiques que subit un individu de sa conception jusqu'à la fin de sa vie.

17 La naissance de l'épidémiologie analytique a suivi la transition épidémiologique dans les pays industrialisés. Comme vous pouvez le voir sur le graphique 1, la mortalité par maladies infectieuses aux États-Unis a chuté au XX^{e} siècle, passant de 800 au début du siècle à 50 pour 100 000 par an en 1980, avant une légère hausse liée au sida[6]. Notez au passage le pic de mortalité associé à la grippe espagnole en 1918-1919. Cette baisse a été

à l'origine d'un bond sans précédent de l'espérance de vie dans l'histoire de l'humanité : vingt-trois années de vie gagnées lors de la première moitié du XX^e siècle. On doit cette baisse avant tout aux progrès de l'hygiène et de l'alimentation, la vaccination et les antibiotiques prenant le relais lors de la seconde moitié du XX^e siècle pour consolider ces résultats.

Graphique 1



Taux de mortalité toutes causes, causes non infectieuses et causes infectieuses, États-Unis, XX^e siècle.

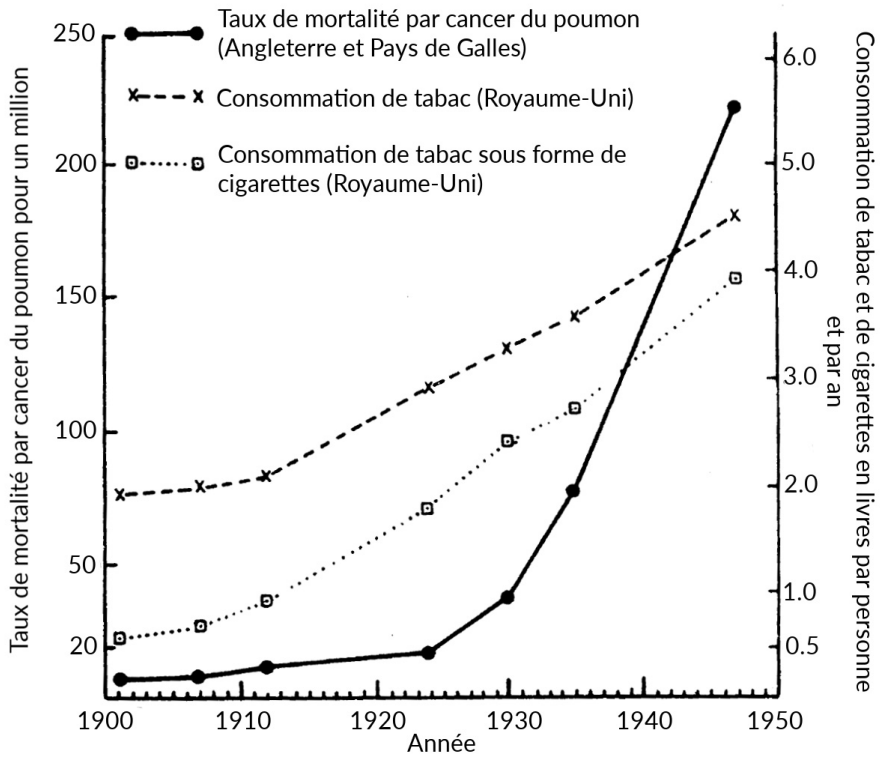
G. L. Amstrong *et al.*, « Trends in infectious disease mortality in the United States during the 20th century », *JAMA*, vol. 281, n° 1, 1999, p. 61-66.

18 Avec l'augmentation de l'espérance de vie, des maladies au développement plus lent allaient prendre le relais des maladies infectieuses comme première cause de mortalité dans les pays industrialisés : ce seront les maladies chroniques, dites non transmissibles, comme le cancer et les maladies cardiovasculaires, puis plus récemment les maladies neurodégénératives. Avec la montée en puissance des maladies dites chroniques, de nouveaux problèmes méthodologiques se profilaient pour l'étude des déterminants des maladies : le modèle – à chaque infection son microbe – ne fonctionnait plus pour ces maladies chroniques d'origine plurifactorielle. Il fallait donc inventer des schémas d'étude et des méthodes analytiques capables d'étudier le rôle étiologique de plusieurs expositions simultanément, et éventuellement leurs interactions. C'est ainsi qu'est née au lendemain de la Seconde Guerre mondiale l'épidémiologie dite « moderne ».

19 Une des premières questions pour les épidémiologistes a été d'étudier le rôle du tabagisme dans l'augmentation de l'incidence du cancer du poumon, en nette augmentation pendant la première moitié du XX^e siècle[7] (graphique 2). La façon la plus simple d'étudier cette question aurait été de réaliser un essai randomisé où la moitié des participants tirés au sort auraient été invités à fumer un paquet de cigarettes par jour, tandis qu'une autre moitié aurait fumé des cigarettes contenant un placebo du

tabac. Vous comprenez sans difficulté les limites d'une telle approche. Il a donc fallu envisager d'autres schémas d'étude, permettant de mettre en évidence les associations entre expositions et maladies, sans intervention de l'investigateur. Ce sont les études observationnelles, comme les études cas-témoins et les études de cohorte.

Graphique 2



Consommation de tabac et de cigarettes (Royaume-Uni) et mortalité par cancer du poumon (Angleterre et Pays de Galles), 1900-1950.

R. Doll et A. B. Hill, « Smoking and carcinoma of the lung », *British Medical Journal*, vol. 2, 1950, p. 739-748.

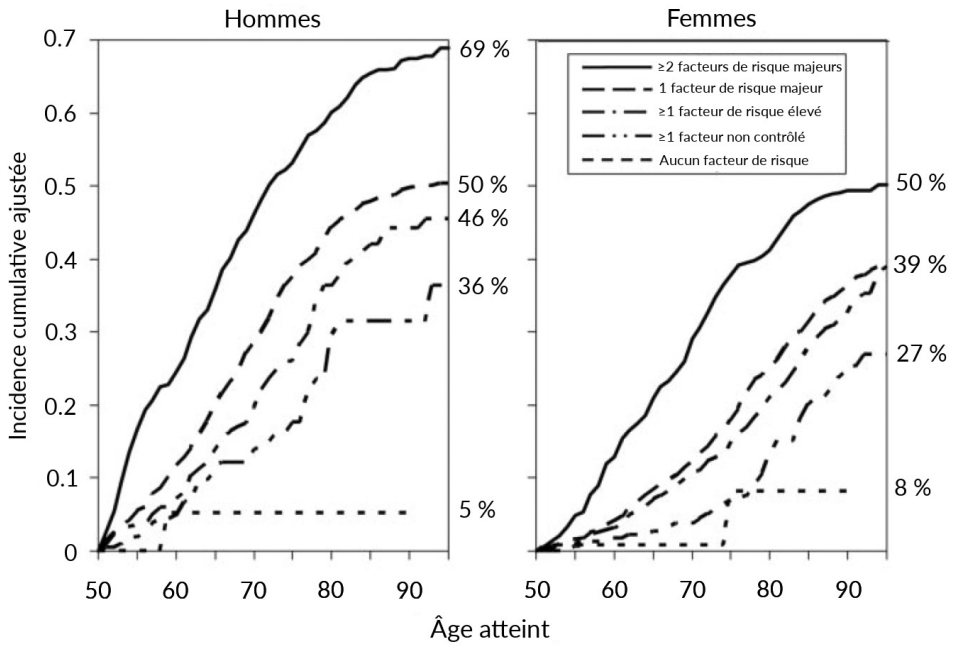
20 La première étude dite « cas-témoins », que nous devons à Richard Doll et Austin Bradford Hill, a vu le jour dans les hôpitaux de Londres en 1948. L'idée était de comparer la consommation tabagique (l'exposition) entre une série de cas, les cancers du poumon (n=709) et les témoins, des patients du même âge et du même sexe, hospitalisés pour des affections non cancéreuses (n=709). Une consommation tabagique supérieure chez les cas comparés aux témoins serait en faveur du rôle délétère du tabac dans la genèse du cancer du poumon. Les résultats de cette étude, publiés dans le *British Medical Journal* en 1950, ont montré en effet une consommation tabagique plus élevée chez les cas comparés aux témoins. On notera que 96 % des témoins hommes fumaient, attestant ainsi de la prévalence très élevée du tabagisme dans la population à l'époque. Chez les femmes, les fumeuses représentaient près de la moitié des participantes chez les témoins.

21 Le deuxième grand sujet d'étude a été celui des facteurs de risques cardiovasculaires. Les États-Unis l'ont réalisé brutalement avec le décès de Franklin D. Roosevelt d'une hémorragie cérébrale en 1945. Ce décès a joué un rôle d'électrochocs parmi les pouvoirs publics aux États-Unis. Harry Truman signe en 1948 le National Heart Act, et octroie 500 000 $ pour débuter une étude de cohorte dédiée aux facteurs de risques

cardiovasculaires dans la ville de Framingham dans le Massachusetts. L'idée est de suivre pendant des dizaines d'années plus de 5000 personnes indemnes de maladies cardiovasculaires au début de l'étude et d'identifier les facteurs associés avec la survenue de maladies coronariennes ou d'accidents vasculaires cérébraux pendant le suivi. C'est ce que nous appelons une « étude de cohorte », par analogie avec les cohortes de soldats romains, qui, une fois enrôlés, restaient à vie dans la cohorte.

22 Cette étude va ainsi identifier plusieurs facteurs de risque – comme, entre autres, l'hypertension artérielle, le tabagisme, l'hypercholestérolémie ou le diabète – associés aux maladies cardiovasculaires. Elle permettra d'estimer le risque cumulatif de développer une maladie cardiovasculaire jusqu'à l'âge de 95 ans selon la présence de ces facteurs de risque à l'âge de 50 ans[8] (graphique 3). En l'absence de facteurs de risque, le risque cumulatif est de seulement 5 % chez les hommes et de 8 % chez les femmes. Il monte à 36 % chez les hommes et à 27 % chez les femmes, avec la présence d'un seul facteur de risque non parfaitement maîtrisé. Cette étude a servi de base à la construction des scores de risque cardiovasculaire utilisés aujourd'hui par les médecins traitants pour décider de la prise en charge des patients.

Graphique 3



Incidence cumulative de maladies cardiovasculaires selon le nombre de facteurs de risques (hypercholestérolémie, hypertension artérielle, diabète, tabagisme) présents à l'âge de 50 ans.

D. M. Lloyd-Jones *et al.*, « Prediction of lifetime risk for cardiovascular disease by risk factor burden at 50 years of age », *Circulation*, vol. 113, 2006, p. 791-798.

23 Les deux schémas d'étude que nous venons de voir, l'étude cas-témoins et l'étude de cohorte, vont s'imposer comme les approches privilégiées pour explorer les associations entre expositions et maladies en épidémiologie observationnelle. L'étape suivante consistera à quantifier l'association entre exposition et maladie, en estimant le risque relatif, c'est-à-dire l'augmentation du risque d'être malade pour les exposés comparés aux non-exposés (dans le cas d'une exposition délétère). Dans une étude de cohorte, il s'agira du ratio entre le taux d'incidence de la maladie chez les exposés et

celui chez les non-exposés. Dans une étude cas-témoins, on utilisera le ratio des cotes (*odds ratio* en anglais), qui correspond mathématiquement au ratio des taux d'incidence de la cohorte virtuelle échantillonnée lors de la réalisation de l'étude cas-témoins. Les techniques d'analyse disponibles aujourd'hui permettent de prendre en compte le fait qu'un individu peut changer de statut d'exposition au cours du temps, ou qu'un temps de latence existe entre exposition et développement de la maladie.

24 Les développements méthodologiques en épidémiologie vont se poursuivre avec une analyse approfondie des biais susceptibles d'être rencontrés lors de la réalisation des études. Vont être décrits des biais de classement, correspondant à des erreurs de mesure dans l'exposition ou la maladie, des biais de sélection, lorsque la relation exposition-maladie est différente dans la population échantillonnée et la population source qui a été échantillonnée, et des biais de confusion. Je vais m'attarder un peu plus sur les biais de confusion qui pourraient vous être moins familiers. Le plus simple est de prendre un exemple. Vous pourriez, lors d'une étude épidémiologique, constater que les individus buveurs de café ont un risque accru de développer un cancer du poumon. Ce risque accru est très vraisemblablement lié à un « effet de confusion » dû au tabagisme : en effet, les buveurs de café fument plus que les non-buveurs, et les fumeurs ont un risque accru de cancer du poumon. Dès lors, l'augmentation du risque de cancer observée chez les buveurs de café est due à la surreprésentation des fumeurs parmi eux, les fumeurs étant davantage exposés au risque de développer un cancer du poumon. C'est ce qu'on appelle un « biais de confusion ». Un facteur de confusion est donc une exposition liée à la fois à l'exposition et à la maladie que vous étudiez. Il pourra ainsi artificiellement augmenter ou diminuer la force de l'association entre exposition et maladie. Les biais de confusion ont fait l'objet de nombreux développements méthodologiques de la part des épidémiologistes pour mieux les prévenir ou les analyser. C'est également un des domaines où la collaboration avec les biostatisticiens a été la plus fructueuse, comme le montrent notamment les modèles d'analyse multivariée de type régression logistique pour les études cas-témoins, ou modèles de Cox pour les études de cohorte, permettant la prise en compte simultanée de plusieurs biais de confusion.

25 Une autre difficulté rencontrée par les épidémiologistes a été la prise en compte des fluctuations aléatoires d'échantillonnage. Nos études sont réalisées sur des échantillons, et donc soumises à des fluctuations aléatoires lors du processus d'échantillonnage. Pour quantifier l'éventualité que les différences observées entre exposés et non-exposés soient dues à des fluctuations aléatoires d'échantillonnage, les épidémiologistes se placent sous l'hypothèse nulle d'absence de différence de risque de maladie entre exposés et non-exposés, et estiment la probabilité d'observer une différence au moins aussi large que celle qui a été observée simplement du fait des variations aléatoires d'échantillonnage : cette probabilité est la valeur P. Si cette probabilité est faible, l'hypothèse nulle d'absence de différence entre exposés et non-exposés est peu plausible, et on conclut qu'il existe une différence de risque entre exposés et non-exposés. Mais quel seuil de la valeur P va-t-on considérer comme suffisamment bas pour rejeter l'hypothèse nulle ? Le choix de ce seuil va avoir un impact considérable sur la production scientifique en recherche épidémiologique et biomédicale : la communauté scientifique et les revues qui s'y rattachent vont utiliser ce seuil comme un couperet pour valider ou non une découverte. Ronald Fisher, bien malgré lui, a été à l'origine du choix de ce seuil. Il est un des plus grands statisticiens du

XX[e] siècle, et a notamment contribué aux approches statistiques permettant d'estimer la valeur P selon la nature des données disponibles. Il considérait lui-même que la valeur P devait être un critère de décision parmi d'autres, et qu'un seuil unique de décision (acceptation ou rejet de l'hypothèse nulle) n'avait pas beaucoup de sens. Il s'est pourtant laissé aller à suggérer qu'une valeur de 0,05 pourrait servir d'aide à la décision[9]. Que n'avait-il pas fait en proposant cette valeur-seuil ? En effet, cette valeur a depuis été reprise par la très grande majorité des revues biomédicales pour différencier des résultats dits « statistiquement significatifs » des autres. Au-dessus de 0,05, vos résultats ne sont plus « significatifs » et ne trouveront pas grâce aux yeux des éditeurs des revues scientifiques. En dessous, votre résultat est validé. Jerzy Neyman et Egon Pearson compléteront la démarche en proposant de chiffrer l'hypothèse alternative, à savoir la différence de risque entre exposés et non-exposés qui nous semble cliniquement ou épidémiologiquement pertinente, afin de permettre le calcul de la taille d'échantillon qui, dès lors, tiendra compte des deux risques assumés par l'investigateur : d'une part, le risque alpha, fixé à 5 %, de conclure à une différence alors qu'il n'y en a pas, et, d'autre part, le risque bêta, fixé communément à 10 % ou 20 %, de ne pas conclure à une différence entre exposés et non-exposés alors qu'elle existe.

26 Alors que les premières études épidémiologiques voyaient le jour, ayant pour finalité d'identifier les déterminants des maladies, on peut s'interroger sur la façon dont cette nouvelle discipline allait s'inscrire dans la démarche scientifique plus générale portant sur la causalité. Un peu à la manière des postulats du microbiologiste Robert Koch publiés en 1890, Austin Bradford Hill a, dès 1965, proposé une série de critères en faveur d'une relation causale entre exposition et maladie[10]. De tous ces critères, un seul est indiscuté, celui de la temporalité : l'exposition doit précéder la survenue de la maladie. Les autres ont fait couler beaucoup d'encre. Un commentaire pour commencer sur la force de l'association : ce critère peut être facilement pris en défaut. Il est cependant intéressant, car il mettait en avant dès cette époque l'importance de la prise en compte de l'amplitude du risque relatif caractérisant l'association entre exposition et maladie, et non pas seulement son caractère statistiquement significatif ou non. Les deux critères qui, à l'usage, me paraissent les plus pertinents sont la reproductibilité et la plausibilité. La reproductibilité fait référence à la réplication des mêmes résultats lors de plusieurs études menées par des observateurs différents : un résultat confirmé par trois études a plus de poids. Même si ce critère n'est pas nécessaire – certains cofacteurs, non ubiquitaires, pourraient être requis pour la matérialisation de la relation causale –, il n'est pas suffisant non plus – les mêmes biais, notamment de confusion, pourraient être présents dans toutes les études. L'autre critère qui me paraît important est la plausibilité, ici le plus souvent de nature biologique. Il est en effet difficile de croire à une relation causale entre exposition et maladie si l'on n'est pas en mesure de décrire un mécanisme biologique qui sous-tend la relation mise en évidence par l'étude épidémiologique. Les épidémiologistes ont, très tôt, pris le parti de s'appuyer sur une autre discipline scientifique pour étayer une relation causale, ce qu'on pourrait appeler le « croisement des preuves », ou ce que Marcus R. Munafò et George Davey Smith ont décrit récemment dans un commentaire de la revue *Nature* comme la « triangulation » des approches[11].

27 Pour illustrer cette question relative à la plausibilité biologique, je vais prendre un exemple parmi mes travaux de recherche sur l'épidémiologie du virus Zika. Vous avez tous en mémoire ces photos terribles d'enfants brésiliens souffrant de microcéphalie.

Le 1er février 2016, suite à la survenue de *clusters* de cas de paralysie périphérique appelée « syndrome de Guillain-Barré » (SGB), et d'anomalies neurologiques congénitales au décours des épidémies de virus Zika, l'Organisation mondiale de la santé (OMS) a décrété une « urgence de santé publique de portée internationale ». Ce faisant, elle demandait à la communauté scientifique de se mobiliser pour étudier le lien qui pouvait exister entre le virus Zika et ces complications dramatiques. Le virus Zika était alors très peu connu. Tout au plus était-il considéré comme une curiosité exotique par les spécialistes de maladies infectieuses et tropicales. Décrit pour la première fois en Afrique en 1947, il avait gagné l'Asie dans les années 1970, et les îles du Pacifique dans les années 2000. La première épidémie d'ampleur a eu lieu en octobre 2013 en Polynésie française, avant une épidémie massive sur le continent latino-américain en 2015.

28 Alors que nous travaillions avec nos collègues de Polynésie française sur les SGB, les premières descriptions de cas de microcéphalie nous parvenaient du Brésil. Nos collègues de Polynésie française avaient bien noté quelques anomalies neurologiques congénitales au décours de l'épidémie, mais pas de foyers de microcéphalie. Nous sommes donc retournés examiner les registres anténataux, notamment d'interruptions de grossesse à visée médicale, et les registres de naissances. Nous avons ainsi pu dénombrer sept cas de microcéphalie en période épidémique, contre un cas hors épidémie. Avec la collaboration de Simon Cauchemez, spécialiste de modélisation mathématique des maladies infectieuses à l'Institut Pasteur, il a été possible de montrer que le modèle qui reproduisait le mieux les données observées était un modèle où le risque de microcéphalie était à son maximum en cas d'infection au premier trimestre de la grossesse, et que le risque était alors de 1 %[12].

29 Parallèlement, les biologistes allaient tout d'abord montrer la plus grande permissivité à l'infection par le virus Zika des cellules progénitrices neurales[13]. Ce sont ces cellules, entre la sixième et la dizième semaine de grossesse, qui initient le développement du cerveau du bébé. Leur atteinte sélective permet de comprendre pourquoi les infections du premier trimestre allaient être celles responsables du risque le plus élevé de malformations congénitales. Ont suivi des études sur des organoïdes, reconstitutions *in vitro* de mini-cerveaux humains à partir de cellules souches pluripotentes, montrant le moindre développement de ces organoïdes après infection par le virus Zika[14]. Et enfin, des modèles animaux, murins et macaques, reproduisant l'infection humaine, ont permis de confirmer ces hypothèses et de tester des vaccins maintenant en cours d'expérimentation chez l'homme.

30 Pendant ce temps, nous poursuivions nos travaux aux Antilles, en collaboration avec les équipes de l'Inserm menées par Bruno Hoen. Là, le suivi d'une cohorte de 546 femmes enceintes atteintes d'une infection symptomatique prouvée par le virus Zika a permis d'estimer à 7 % le risque d'anomalies congénitales chez les enfants nés de mères infectées pendant la grossesse[15]. Ce risque était maximal, à 12,7 %, pour des infections contractées au cours du premier trimestre de la grossesse. La comparaison avec une série de femmes non exposées, c'est-à-dire n'ayant pas été infectées par le virus Zika, nous a permis depuis de réduire cette estimation à un risque de 4,1 % d'anomalies congénitales sévères pour les infections du premier trimestre réellement attribuables au virus Zika. L'intérêt de cet exemple est de montrer comment les études épidémiologiques et biologiques se renseignent mutuellement pour aboutir à l'établissement d'une relation de causalité, ici entre le virus Zika et les anomalies

neurologiques congénitales chez les fœtus et les bébés. L'apport de l'épidémiologie est non seulement de montrer que les processus mécanistiques révélés par les études biologiques ont un impact clinique en population humaine, mais également de quantifier cet impact.

Grandes contributions de l'épidémiologie à la santé publique

31 Maintenant que nous comprenons mieux les méthodes utilisées en épidémiologie et comment sont identifiés les facteurs responsables de maladies, il est temps d'examiner les grandes contributions de l'épidémiologie à la santé publique.

32 Nous avons parlé de la cohorte de Framingham qui a mis en évidence les facteurs de risque cardiovasculaire. Une autre cohorte américaine, la Nurses' Health Study, s'est attachée à décrire les conséquences sur la santé de la prise d'hormones – de type œstrogène et progestérone – et des comportements alimentaires. Les effectifs sont considérables : plus de 200 000 femmes suivies, pour certaines depuis plus de quarante ans.

33 Cette cohorte a produit un nombre très élevé de publications scientifiques qu'il ne serait pas possible de résumer en quelques lignes. On peut citer six recommandations issues de ces études permettant de réduire de 50 % le risque de développer un cancer chez une femme : ne pas fumer ; pratiquer une activité physique régulière ; avoir un poids raisonnable ; avoir une nourriture riche en fruits, légumes, céréales complètes, fibres alimentaires, et pauvre en acides gras saturés et trans ; prendre des suppléments multivitaminés quotidiennement ; et limiter la durée du traitement hormonal post-ménopause[16]. Ces recommandations sont maintenant connues de tous, et c'est une très bonne chose. Ce que beaucoup ne savent pas, c'est qu'elles sont issues d'études épidémiologiques portant sur des centaines de milliers d'individus suivis pendant plusieurs dizaines d'années. Cette cohorte nous a également apporté une des preuves les plus solides de la toxicité des particules fines[17], en montrant le lien entre la concentration moyenne de particules fines des douze derniers mois mesurée sur une échelle de 6 km^2 et la mortalité des infirmières de la cohorte sur une période de sept ans.

34 Les cohortes que nous venons de décrire sont des études épidémiologiques réputées fiables. Elles sont cependant très longues et coûteuses. Les études cas-témoins ont moins bonne réputation dans le monde des épidémiologistes, du fait des risques de biais de sélection, notamment pour les témoins, et de la moins bonne mesure des expositions qui se fait de façon rétrospective. Néanmoins, il est des circonstances où les études cas-témoins ont pu rapidement apporter des éléments probants en faveur d'une association, quitte à ce que ces premiers résultats soient confirmés par la suite au cours d'études de cohorte.

35 Une étude cas-témoins emblématique a été menée en Grèce par Dimitri Trichopoulos, la première à documenter le risque de cancer du poumon associé au tabagisme passif[18]. L'étude a tiré parti du fait que les femmes fumaient très peu dans les années 1970 en Grèce. Leur exposition au tabac était donc essentiellement liée au tabagisme de leur mari. L'étude a donc comparé la consommation tabagique des maris de femmes atteintes d'un cancer du poumon à celle de femmes sans cancer du poumon.

L'augmentation du risque de cancer du poumon lié au tabagisme passif était de l'ordre de 2 à 3 selon le niveau de tabagisme du mari. Cette étude a eu un impact considérable en santé publique en donnant une justification sanitaire à l'interdiction du tabagisme dans les lieux publics, puisqu'un fumeur ne mettait plus simplement sa vie en danger en fumant, mais celle des autres. Les implications légales et financières de cette étude ont été considérables, comme vous vous en doutez.

36 Un autre exemple pour lequel les études cas-témoins ont joué un rôle déterminant est celui de la mort subite du nourrisson. Beaucoup d'entre vous se souviennent de la crainte suscitée par cette affection, qui touchait prioritairement les nourrissons âgés de deux à quatre mois, préférentiellement les garçons. À la suite d'une série d'études cas-témoins, 19 au total, il a été montré que le coucher en position ventrale était un facteur de risque important de la mort subite du nourrisson[19]. Ainsi, les recommandations en faveur du coucher dorsal ont permis, en France, de réduire le nombre de cas de 70 % en cinq ans (1464 décès en 1991 contre 451 en 1996).

La période du doute en épidémiologie

37 Après ce que j'appellerais les « Trente glorieuses de l'épidémiologie », de 1960 à 1990, pendant lesquelles les principaux facteurs de risque des maladies cardiovasculaires et de nombreux cancers ont été identifiés, s'installe une période de doute au début des années 1990.

38 Les critiques les plus vives viendront de la plume de Petr Skrabanek, un professeur iconoclaste du Trinity College à Dublin. Petr Skrabanek dénonça avec vigueur les « *riskfactorologists* » qui, à partir d'études de cohorte de grande taille, ont établi des listes de comportements à risque sur la base d'associations statistiques dénuées de fondement biologique ou de pertinence clinique. Les messages de prévention qui en découlèrent, emprunts de moralisme (ne buvez pas d'alcool, ne fumez pas, ne multipliez pas les partenaires sexuels, etc.), nous condamnent, selon Skrabanek, à mener de longues vies d'ascètes.

39 Skrabanek met en cause le croisement, sans modèle biologique sous-jacent, de toutes les données disponibles concernant les expositions et les maladies dans les cohortes de grande taille, rapportant des associations peu plausibles qui furent reprises sans discernement par la presse médicale et les journaux grand public, comme par exemple l'effet supposément protecteur de la consommation de tomates contre le cancer du col de l'utérus, ou l'augmentation du risque de fausses couches chez les femmes enceintes qui consomment du café, mais seulement si elles ont des nausées matinales[20].

40 Une analyse quantitative argumentée de la *riskfactorology* chère à Petr Skrabanek vint de l'article de John Ioannidis au titre provocateur paru dans *PLoS Medicine* en 2005 : « Why most published research findings are false[21] ». Dans cet article, il montre que la majorité des associations dites « statistiquement significatives » publiées dans les revues biomédicales sont des associations dues à des fluctuations aléatoires d'échantillonnage, pour peu que le nombre d'associations testées « pour voir », pour l'immense majorité non existantes, surpasse très largement le nombre d'associations réelles testées lors de l'enquête. Ce phénomène est par ailleurs amplifié par l'attitude des revues biomédicales qui privilégient la publication de résultats dits « significatifs » au détriment des résultats négatifs pourtant issus d'études de qualité. Parmi les solutions proposées, on retrouve la pratique de pré-spécifier dans les protocoles de

recherche un nombre restreint d'hypothèses à tester, l'utilisation de seuils de valeur P plus rigoureux pour parler de significativité – en génétique, discipline dans laquelle des millions de tests peuvent être pratiqués, le seuil de significativité de la valeur P a été abaissé à $5*10^{-8}$ –, et un travail auprès des éditeurs destiné à les convaincre de l'importance des résultats non significatifs lorsqu'ils proviennent d'études bien menées.

41 Plus inquiétante, car moins facile à débusquer, et remettant en cause la validité des études de cohorte dans leur ensemble, a été la non-confirmation par des essais randomisés de la protection contre les maladies cardio-vasculaires conférée par les bêta-carotènes[22]. Les bêta-carotènes sont présents dans les fruits et les légumes. Dans une série d'études de cohorte américaines et européennes, les bêta-carotènes semblaient diminuer le risque cardio-vasculaire. Ces résultats étaient d'autant plus acceptés que les propriétés anti-oxydantes des bêta-carotènes pouvaient expliquer qu'ils préviennent l'athérosclérose. Comme les bêta-carotènes peuvent être pris sous forme de compléments alimentaires, il a été possible de les administrer contre placebo au cours d'essais randomisés entre deux groupes de participants rendus identiques par le tirage au sort ; et l'effet protecteur des études de cohorte n'a pas été retrouvé. On a même observé un effet délétère d'amplitude modérée. Ceci nous amène à remettre en cause les résultats des études de cohorte, toutes atteintes du même biais de confusion, lié à des comportements protecteurs contre le risque cardiovasculaire associés à la prise de bêta-carotène. On peut supputer que les grands consommateurs de bêta-carotène faisaient partie de ces personnalités soucieuses de leur santé (*health conscious* en anglais), qui ont un mode de vie associant toutes les bonnes pratiques en termes de prévention du risque cardiovasculaire, c'est-à-dire ne fumant pas, faisant de l'exercice, ayant une alimentation équilibrée et riche en fruits et légumes, surveillant leur tension artérielle et leur cholestérol, etc. Ces bonnes pratiques, associées à la consommation de bêta-carotène, exerçaient donc un effet dit « de confusion » sur l'association entre bêta-carotène et risque cardiovasculaire, que les méthodes d'ajustement statistique par analyse multivariée ne parvenaient pas à contrôler parfaitement dans les études de cohorte. La difficulté à gérer ce type de biais de confusion a entraîné de nouvelles approches méthodologiques. On peut citer notamment la randomisation mendélienne.

42 La randomisation mendélienne s'appuie sur la propriété suivante des variants génétiques : étant distribués de façon aléatoire lors de la conception, ils ne peuvent être associés à des comportements qui pourraient jouer le rôle de facteur de confusion dans les études épidémiologiques. Si les variants étudiés reproduisent les effets d'un comportement particulier, par exemple une addiction à l'alcool ou au tabac, ils permettent d'étudier l'effet de ces comportements comme le ferait une étude randomisée. L'exemple suivant montre l'intérêt d'une telle approche. De nombreuses études de cohorte ont montré qu'une consommation modérée d'alcool, par exemple un à deux verres de vin par jour, est associée à une mortalité plus faible que l'abstinence, et ce en partie du fait d'une mortalité cardiovasculaire moins élevée. Plusieurs hypothèses ont été évoquées : 1) qu'il existe une explication biologique réelle, notamment une élévation du « bon » cholestérol, les HDL, chez les buveurs de vin ; 2) qu'il y ait une proportion non négligeable de personnes en mauvaise santé parmi les abstinents, à qui l'on a dit d'arrêter de boire du fait de leur mauvaise santé ; 3) que les buveurs modérés correspondent à des personnalités « *health conscious* » précédemment citées, qui contrôlent bien leurs habitudes et leur mode de vie. Pour essayer de comprendre mieux ce phénomène, des épidémiologistes ont étudié les porteurs d'un

variant du gène ADH1B codant pour les enzymes alcool-déshydrogénases associées à la dépendance à l'alcool. Les porteurs du variant ont une consommation d'alcool plus faible, moins de dépendance à l'alcool et, dans cette étude, un risque plus faible d'infarctus du myocarde[23]. Il semblerait donc que même les buveurs modérés d'alcool gagneraient à moins boire. Petr Skrabanek va se retourner une fois de plus dans sa tombe.

43 Une autre limite de l'épidémiologie s'est manifestée lors de la prise en compte des expositions environnementales. Ces expositions sont des agents physiques, chimiques et biologiques présents dans l'air, l'eau, les sols ou l'alimentation. Les exemples sont nombreux : particules fines dans l'atmosphère, produits phytosanitaires, métaux lourds, etc., et représentent des sujets de société importants. Dans ce contexte, les expositions individuelles sont très difficiles à mesurer. Alors qu'il est relativement facile d'estimer la consommation tabagique d'un individu au cours des vingt dernières années avec un questionnaire, c'est une tout autre affaire que d'estimer son exposition à un contaminant présent dans l'air, l'eau ou l'alimentation. De plus, ces expositions sont multiples et souvent simultanées : les sujets sont le plus souvent exposés à plusieurs insecticides en même temps. L'un d'entre eux peut être responsable d'un effet sanitaire, et les autres être sans effet. Comment savoir lequel est incriminé ? Comment faire la part des effets synergiques qui peuvent exister entre plusieurs insecticides ? Comment prendre en compte le temps de latence entre l'exposition et le développement de la maladie ? Par exemple, dans le cas des perturbateurs endocriniens, c'est une exposition *in utero* qui peut être responsable d'un effet à l'âge adulte. Enfin, la plupart des effets liés aux expositions environnementales est faible, et il faut donc des tailles d'échantillon importantes pour les mettre en évidence. Cependant, comme beaucoup de sujets sont exposés, l'impact en santé publique peut être important. Face à ces difficultés, l'épidémiologie atteint ses limites, notamment pour répondre à la demande des pouvoirs publics, qui veut une garantie que le produit est sans danger pour la santé. Les études les plus probantes ont été celles menées chez des sujets surexposés au cours de leur pratique professionnelle, mais pour l'estimation du risque en population générale, l'avenir est aux cohortes de grande taille, certaines d'entre elles démarrant dès la vie fœtale, et utilisant des marqueurs biologiques ou des capteurs de pollution. Dans l'état actuel des choses, en l'absence d'effet documenté chez l'homme, la réponse est politique : c'est le principe de précaution face à des produits ayant démontré une toxicité *in vitro* ou chez l'animal, mais pour lesquels un effet chez l'homme n'a pas été mis en évidence.

44 Enfin, l'épidémiologie est aujourd'hui sous tension, à la croisée de plusieurs chemins. Doit-elle se focaliser sur les déterminants génétiques des maladies, répondant ainsi aux possibilités offertes par les nouvelles technologies de séquençage du génome humain ? Doit-elle au contraire garder le cap sur les déterminants comportementaux, addictifs et alimentaires, dont on sait qu'ils continuent de représenter la part la plus importante et modifiable des causes de nos maladies ? Ou doit-elle embrasser de façon plus complète l'ensemble des expositions, regroupé sous le vocable d'*exposome*, et prendre en compte les dimensions sociales et communautaires contribuant au développement de nos états de santé ? À titre d'exemple, on observe sur la période 1976-2013 en France une différence stable de trois ans d'espérance de vie à 35 ans entre femmes cadres et ouvrières, et de six ans et demi entre hommes cadres et ouvriers[24]. À quoi sont dus ces écarts d'espérance de vie ? S'agit-il de différences de comportements, d'expositions environnementales, d'accès aux soins, ou de qualité de soins, entre cadres et ouvriers ?

Il est vrai que les épidémiologistes ne sont pas très à l'aise quand il s'agit de parler d'inégalités sociales et des moyens d'y remédier. Une source d'inspiration nous vient du modèle des boîtes chinoises imbriquées (comprenez « poupées russes ») de Mervyn Susser, chaque boîte représentant un niveau emboîté dans le niveau supérieur, allant du moléculaire jusqu'au social, les modèles statistiques dits « multiniveaux » permettant d'appréhender l'ensemble de ces composantes au sein d'une même analyse[25]. Une autre approche est celle proposée par Merrill Singer, un anthropologue américain qui a créé le concept de « *syndemics* » ou « *synergistic epidemics* » pour décrire le regroupement et l'interaction de co-morbidités (par ex. : surpoids, diabète, dépression) chez les plus vulnérables, et la façon avec laquelle ces regroupements vont être entretenus par un contexte social, environnemental ou économique défavorable[26].

L'avenir de l'épidémiologie

45 Je terminerai cette leçon inaugurale par quelques réflexions sur l'avenir de l'épidémiologie. Comme c'est le cas pour beaucoup d'autres disciplines scientifiques, l'avenir de l'épidémiologie va être profondément touché par l'irruption des analyses des *big data*. Les *big data*, ou données massives, nous dépassent par leur abondance et leur diversité. Le premier défi va être de recueillir des données de natures très différentes, comme celles du séquençage du génome, mais également celles des objets connectés, des dossiers médicaux électroniques, de l'imagerie médicale, des réseaux sociaux ou des capteurs de pollution atmosphérique. Il faudra ensuite vérifier ces données, les transformer, les stocker, les partager, les analyser, et enfin en visualiser les résultats d'analyse, le tout dans un contexte réglementaire de plus en plus strict, qui vise à protéger les données des personnes.

46 Une des initiatives les plus avancées dans le domaine est portée par la cohorte appelée UK Biobank. Près de 500 000 individus ont été recrutés depuis 2006 au sein de cette étude de cohorte où les données médicales sont mises en relation avec celles du séquençage du génome. Au-delà de l'effort gigantesque que représente la constitution de cette cohorte, il faut noter la volonté des investigateurs et des financeurs de rendre les données accessibles à la communauté scientifique. Tout chercheur peut demander à y avoir accès, et cet accès sera autorisé après évaluation scientifique de la proposition. D'autres cohortes, américaines pour la plupart, sont en cours de constitution dans les milieux académiques et privés, et portent sur des centaines de milliers d'individus également. On peut citer la Million Veteran Study et le All of Us Research Program. Il peut s'agir également de *consortia* où sont réunies plusieurs initiatives nationales portant sur les mêmes thématiques de recherche, comme les mesures anthropométriques (taille et poids) ou les maladies psychiatriques.

47 Où se situe la France dans cette compétition internationale ? Historiquement, la première cohorte française fut la cohorte Gazel, qui suivait les employés d'EDF/GDF. Elle s'est penchée sur la santé au travail et sur les déterminants sociaux de la santé. Depuis, plusieurs cohortes ont vu le jour. On peut citer, entre autres, les cohortes proposant un suivi sur internet, NutriNet-Santé, sur les liens entre alimentation et santé, et la cohorte ComPaRe, pour les patients atteints de maladies chroniques. Il y a également la cohorte Constances, dont l'originalité est de s'appuyer sur un échantillonnage représentatif de la population française. Il y a enfin plusieurs cohortes pédiatriques (Eden et Elfe) et thématiques (notamment sur le VIH et les hépatites

virales). La génomique est encore peu présente dans ces cohortes, mais elle pourrait le devenir en interface avec le Plan France Génomique 2025, dont l'objectif est de mettre à disposition des plateformes pour le séquençage à visée de diagnostic médical.

48 Que pouvons-nous dire des résultats de ces études de cohorte ? Après l'enthousiasme des débuts, marqué par l'espoir de prédire notre risque individuel de développer telle ou telle maladie sur la base de nos données génétiques, est venue une phase de perplexité. Certes, les variants génétiques associés aux maladies communes qui ont été récemment identifiés se comptent par milliers, mais l'augmentation de risque associée à chacun de ces variants est faible, et leur action est conditionnée par la présence d'autres variants qu'il reste à identifier ou de modifications épigénétiques rendant la généralisation des résultats très complexe. Le salut viendra peut-être des scores polygéniques semblables à ceux récemment publiés par l'équipe de Sekar Kathiresan, validés sur les données de la UK Biobank, et offrant une prédiction pour la maladie coronarienne, la fibrillation auriculaire, le diabète, les maladies inflammatoires de l'intestin et le cancer du sein, avec une augmentation de risque conséquente (au moins triplement) pour des scores atteints par une proportion non négligeable de la population (1,5 à 8 % des individus[27]). Les auteurs de cette étude concluent en proposant que chacun teste ses prédispositions génétiques pour les maladies étudiées, afin d'adopter un programme de prévention et de dépistage adapté selon les résultats de ces tests. Il s'agit donc d'un élargissement du concept de médecine de précision à la prévention. La médecine de précision vise en effet à proposer à chaque individu des traitements adaptés à ses propres caractéristiques génétiques, ou à celle de sa tumeur en cas de cancer. Des avancées remarquables ont été faites dans le traitement de certains cancers, mais à des coûts considérables. On peut imaginer que se mette en place demain une médecine de précision axée sur la prévention, qui offrira à chacun des messages de prévention ajustés à sa tolérance génétiquement déterminée à la nicotine ou à l'alcool, adaptée au risque qu'il a de développer un diabète, une hypertension artérielle, etc.

49 C'est d'ailleurs ce que fait le site de 23andMe, une compagnie californienne dont Google est actionnaire, où l'on peut, pour moins de 200 $, effectuer un test génétique à partir de sa salive. Ce test vous donnera des informations sur vos origines ancestrales et sur la présence d'allèles associés à certaines maladies (entre autres, maladie de Parkinson ou d'Alzheimer). Outre les questions d'éthique et de confidentialité des données que je n'aborderai pas ici, il est intéressant de s'interroger sur l'efficacité de l'élargissement du concept de la médecine de précision à la prévention. Tout ceci est assez séduisant sur le papier, mais nous laisse face à de nombreuses incertitudes. D'une part, proposer un dépistage de la maladie de Parkinson ou de la maladie d'Alzheimer, alors qu'il n'existe pas de mesures de prévention réellement établies pour ces maladies, est très anxiogène. D'autre part, nous ne savons pas si les individus sont prêts à changer de comportement une fois informés d'une prédisposition génétique à développer une maladie donnée. L'analyse de dix-huit essais randomisés nous répond par la négative[28]. Nous devrons aussi changer de paradigme : l'approche de santé publique s'est jusqu'alors appuyée sur des messages de prévention universels, adressés à la collectivité, où l'effet d'entraînement a sa place. Nous passerions à un modèle individualiste, plus coûteux, inégalitaire, et vraisemblablement moins performant. La lecture de l'article de Geoffrey Rose « Sick individuals and sick populations » est à ce titre riche d'enseignements[29]. Je suis convaincu qu'une meilleure connaissance des déterminants génétiques des maladies est fondamentale pour les études mécanistiques

qui permettront, selon la fonction des gènes concernés, d'identifier des cibles pour des vaccins ou des traitements. On pourra aussi envisager, pour certaines maladies, des applications de la thérapie génique. Je suis en revanche dubitatif quant à l'utilisation du dépistage des prédispositions génétiques à des fins de prévention individualisée pour les maladies communes. Rappelons-nous que les études ayant démontré le bénéfice du dépistage de certains cancers, comme par exemple le cancer du sein, ont porté sur des centaines de milliers de femmes suivies pendant plus de dix ans, avant de savoir quel type de dépistage proposer et à quel âge.

50 Qu'en est-il des méthodes d'analyse ? Quand on parle de *big data*, l'intelligence artificielle n'est pas loin. Nous savons déjà que l'intelligence artificielle excelle dans les domaines où existent des règles, un nombre fini de solutions, ou des jeux de données pour l'apprentissage. Elle a fait ses preuves dans le diagnostic médical, notamment l'identification des mélanomes et le diagnostic de la rétinopathie diabétique. Dans le domaine de l'identification des gènes associés à des maladies, les méthodes statistiques classiques n'ont pas encore été supplantées. En revanche, dans le domaine de la prédiction, l'intelligence artificielle est de plus en plus performante. S'il s'agit par exemple de prédire la mortalité intra-hospitalière à partir d'une base de données cliniques et biologiques de 40 000 individus réanimés après un arrêt cardiaque, les méthodes d'apprentissage automatique sont plus performantes que les méthodes statistiques classiques – ce sont d'ailleurs ces approches qui sont utilisées en marketing pour vous faire des propositions d'achat sur votre smartphone[30]. On oppose cependant à ces méthodes d'apprentissage automatique le fait qu'elles fonctionnent comme des boîtes noires, ne permettant pas d'identifier les variables explicatives, celles dans notre exemple qui sont associées à la mortalité intra-hospitalière, et que leurs résultats ne sont pas reproductibles. Il est vraisemblable que l'intelligence artificielle supplantera un jour les méthodes statistiques classiques, y compris dans la capacité à fournir ces variables explicatives. Il est un domaine cependant où l'intelligence artificielle est encore prise en défaut : c'est celui de la distinction entre prédiction et inférence causale. De façon triviale, l'utilisation d'insuline prédit très bien la présence d'un diabète, mais elle n'est pas causale. L'étape d'inférence causale est importante, car elle seule permet l'identification des caractéristiques des patients, qui, si elles étaient modifiées en amont, permettraient d'améliorer leur pronostic. Or, dans l'état actuel de notre savoir, seule la connaissance *a priori* des relations existant entre les différentes variables permet de faire la distinction entre prédiction et causalité – rappelez-vous l'exemple de l'insuline et du diabète. C'est ce que proposent les graphes acycliques orientés, nouveau développement méthodologique en épidémiologie permettant d'appréhender l'inférence causale[31]. Et c'est cette connaissance *a priori* des relations entre variables qui fait aujourd'hui défaut aux modèles d'apprentissage automatique.

51 Quelques mots pour terminer ce panorama sur la santé publique de précision, qui est à la santé publique ce que la médecine de précision est à la médecine. Dans les deux cas, on utilise les *big data*, mais pour la santé publique de précision, l'échelle d'analyse est la population, tandis que pour la médecine de précision, l'échelle d'analyse est l'individu. Il s'agit donc d'agréger des données massives médicales, administratives, environnementales, ou provenant des réseaux sociaux, à une échelle populationnelle fine, et de cibler des interventions sur des populations beaucoup mieux caractérisées. On peut citer à titre d'exemple une étude dans laquelle les investigateurs ont regroupé des données de conflits armés en Afrique à une échelle de 100 km^2 sur une période de vingt ans, et ont extrait des enquêtes démographiques de santé des données de

mortalité infantile à la même échelle[32]. L'analyse longitudinale a permis d'estimer l'excès de mortalité infantile pendant les périodes de conflits armés. Les résultats sont édifiants : il y a eu trois fois plus de décès chez les moins de 1 an dus à la privation de soins lors des conflits armés que de décès directement liés aux conflits : 3 millions de décès *versus* 1 million de décès sur une période de vingt ans.

52 Dans le domaine des maladies infectieuses, l'utilisation d'appareils portables pour séquencer le génome des virus, couplée à des systèmes de transmission des données à partir de zones reculées, devrait permettre l'analyse en temps réel des chaînes de transmission de virus épidémiques. Ce type d'analyse a été réalisé *a posteriori* en utilisant les données de séquençage du virus Ebola pendant l'épidémie de 2014-2015 en Afrique de l'Ouest[33]. L'espoir est que ce type de données puisse devenir disponible en temps réel et guider les interventions de contrôle de l'épidémie. Il ne faut pas oublier cependant que les terrains sur lesquels les équipes interviennent aujourd'hui sont très difficiles, et que les priorités du moment sont au renforcement des capacités de surveillance et de réponse dans ces pays aux systèmes de santé très fragiles, avant d'envisager l'utilisation d'outils sophistiqués.

*

53 Le modèle que nous voyons émerger est celui de très grandes cohortes, associant aux données médicales celles du génome et de l'exposome. Ces cohortes sont la promesse de grandes avancées dans l'identification des gènes associés aux maladies et dans la compréhension de l'impact de notre environnement sur notre santé, domaines sur lesquels l'épidémiologie avait buté jusqu'à récemment. L'histoire de la santé publique nous a appris que ce sont les actions de prévention collectives, menées dans le souci de toucher le plus grand nombre, et notamment les plus défavorisés, qui ont permis les plus grands progrès en santé humaine. Ce modèle va inévitablement évoluer vers une connaissance beaucoup plus précise et individualisée des déterminants de notre santé. Il me paraît fondamental de préparer nos sociétés dès maintenant à l'intégration de ce nouveau savoir dans notre quotidien. Le temps viendra où une femme enceinte pourra connaître dès les premières semaines de grossesse, par une simple prise de sang, les caractéristiques génétiques de son enfant à naître. Sinon, un simple prélèvement de salive après la naissance permettra de connaître les prédispositions génétiques de l'enfant pour un nombre incalculable de maladies. Sommes-nous prêts, de façon à nous assurer que ce savoir sera utilisé pour le meilleur, par exemple la thérapie génique ciblée pour les pathologies les plus graves, et non pour le pire, qui pourrait dans ce contexte se rapprocher de l'eugénisme ? Pour les expositions environnementales, nous saurons mieux apprécier les effets délétères de certains produits. La vraie question sera l'appréciation de la balance bénéfice-risque, car l'idée que les produits auxquels nous sommes exposés soient dépourvus de risque est utopique. Il s'agira ensuite de conjuguer nos efforts pour préserver, de concert, le progrès technologique, la santé de l'homme et celle de la planète.

NOTES

1. É. Beaugrand, *L'Hygiène, ou l'art de conserver la santé*, Paris, Librairie Hachette et C^{ie}, 1855.

2. C. E. A. Winslow, « The untilled field of public health », *Modern Medicine*, n° 2, 1920, p. 183-191.

3. *État de santé de la population en France – Rapport 2017*, Drees et Santé publique France, mai 2017 [en ligne : https://drees.solidarites-sante.gouv.fr].

4. A. C. Ghani, N. M. Ferguson, C. A. Donnelly et R. M. Anderson, « Predicted vCJD mortality in Great Britain », *Nature*, vol. 406, 2000, p. 583-584.

5. A.-J. Valleron, P. Y. Boelle, R. Will et J. Y. Cesbron, « Estimation of epidemic size and incubation time based on age characteristics of vCJD in the United Kingdom », *Science*, vol. 294, 2001, p. 1726-1728.

6. G. L. Amstrong, L. A. Conn et R. W. Pinner, « Trends in infectious disease mortality in the United States during the 20th century », *JAMA – Journal of the American Medical Association*, vol. 281, n° 1, 1999, p. 61-66.

7. R. Doll et A. Bradford Hill, « Smoking and carcinoma of the lung », *British Medical Journal*, vol. 2, 1950, p. 739-748.

8. D. M. Lloyd-Jones, E. P. Leip, M.G. Larson, R.B. D'Agostino, A. Beiser, PW Wilson, PA Wolf, D. Levy, « Prediction of lifetime risk for cardiovascular disease by risk factor burden at 50 years of age », *Circulation*, vol. 113, 2006, p. 791-798.

9. R. A. Fisher, *Statistical Methods for Research Workers*, Édimbourg (Tweeddale Court)/ Londres, Oliver & Boyd, 1925, p. 70.

10. A. Bradford Hill, « The environment and disease: association or causation? », *Proceedings of the Royal Society of Medicine*, vol. 58, 1965, p. 295-300.

11. M. R. Munafò et G. Davey Smith, « Repeating experiments is not enough », *Nature*, vol. 553, 2018, p. 399-401.

12. S. Cauchemez, M. Besnard, P. Bompard *et al.*, « Association between Zika virus and microcephaly in French Polynesia, 2013-15: a retrospective study », *Lancet*, vol. 387, n° 10033, 2016, p. 2125-2132.

13. H. Tang, C. Hammack, S. C. Ogden *et al.*, « Zika virus infects human cortical neural progenitors and attenuates their growth », *Cell Stem Cell*, vol. 18, n° 5, 2016, p. 587-590.

14. P. P. Garcez, E. C. Loiola, R. Madeiro da Costa *et al.*, « Zika virus impairs growth in human neurospheres and brain organoids », *Science*, vol. 352, n° 6287, 2016, p. 816-818.

15. B. Hoen, B. Schaub, A. L. Funk *et al.*, « Pregnancy outcomes after Zika infection in the French territories of America », *The New England Journal of Medicine*, vol. 378, n° 11, 2018, p. 985-994.

16. G. A. Colditz et S. E. Hankinson, « The nurses' health study: lifestyle and health among women », *Nature Reviews Cancer*, vol. 5, n° 5, 2005, p. 388-396.

17. J. E. Hart, X. Liao, B. Hong *et al.*, « The association of long-term exposure to PM2.5 on all-cause mortality in the nurses' health study and the impact of measurement-error correction », *Environmental Health*, vol. 14, n° 1, p. 38.

18. D. Trichopoulos, A. Kalandidi, L. Sparros et B. MacMahon, « Lung cancer and passive smoking », *Internal Journal of Cancer*, vol. 27, 1981, p. 1-4.

19. S. M. Beal et C. F. Finch, « An overview of retrospective case-control studies investigating the relationship between prone sleeping position and SIDS », *Journal of Paediatrics and Child Health*, vol. 27, 1991, p. 334-339.

20. P. Skrabanek, « The poverty of epidemiology », *Perspectives in Biology and Medicine*, vol. 35, n° 2, 1992, p. 183.

21. J. P. A. Ioannidis, « Why most published research findings are false », *PLoS Medicine*, vol. 2, n° 8, 2005, e124.

22. M. Egger, M. Schneider et G. Davey Smith, « Spurious precision? Meta-analysis of observational studies », *British Medical Journal*, vol. 316, 1998, p. 140-144.

23. V. H. Holmes, C. E. Dale, L. Zuccolo *et al.*, « Association between alcohol and cardiovascular disease: Mendelian randomization analysis based on individual participant data », *British Medical Journal*, vol. 349, 2014, g4164.

24. *État de santé de la population en France – Rapport 2017*, *op. cit.*, p. 22.

25. M. Susser et E. Susser, « Choosing a future for epidemiology II. From black box to Chinese boxes and eco-epidemiology », *American Journal of Public Health*, vol. 86, n° 5, 1996, p. 674-677.

26. M. Singer, « A dose of drugs, a touch of violence, a case of AIDS: conceptualizing the SAVA syndemic », *Free Inquiry in Creative Sociology*, vol. 24, n° 2, 1996, p. 99-110.

27. A. V. Khera, M. Chaffin, K. G. Aragam *et al.*, « Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations », *Nature Genetics*, vol. 50, 2018, p. 1219-1224.

28. G. J. Hollands, D. P. French, S. J. Griffin *et al.*, « The impact of communicating genetic risks of disease on risk-reducing health behavior: systematic review with meta-analysis », *British Medical Journal*, vol. 352, 2016, i1102.

29. G. Rose, « Sick individuals and sick populations », *International Journal of Epidemiology*, vol. 30, n° 3, 2001, p. 427-432.

30. S. Nanayakkara, S. Fogarty, M. Tremeer *et al.*, « Characterising risk of in-hospital mortality following cardiac arrest using machine learning: A retrospective international registry study », *PLoS Medicine*, vol. 15, n° 11, 2018, e1002709.

31. S. Greenland, J. Pearl et J. M. Robins, « Causal diagrams for epidemiologic research. *Epidemiology* », vol. 10, n° 1, 1999, p. 37-48.

32. Z. Wagner, S. Heft-Neal, Z. A. Bhutta, R. E. Black, M. Burke et E. Bendavid, « Armed conflict and child mortality in Africa: a geospatial analysis », *The Lancet*, vol. 392, n° 10150, 2018, p. 857-865.

33. G. Dudas, L. M. Carvalho, T. Bedford *et al.*, « Virus genomes reveal factors that spread and sustained the Ebola epidemic », *Nature*, vol. 544, 2017, p. 309-315.

AUTEUR

ARNAUD FONTANET

Professeur invité au Collège de France sur la chaire annuelle Santé publique 2018-2019